Personalized Generation of Word Clouds From Tweets

Martin Leginus
Department of Computer Science, Aalborg University, Selma Lagerlöfs Vej 300, Aalborg 9200, Denmark. E-mail: martin.leginus@gmail.com

ChengXiang Zhai
Department of Computer Science, University of Illinois at Urbana-Champaign, 201 North Goodwin Ave, Urbana, IL 61801. E-mail: czhai@illinois.edu

Peter Dolog
Department of Computer Science, Aalborg University, Selma Lagerlöfs Vej 300, Aalborg 9200, Denmark. E-mail: dolog@cs.aau.dk

Active users of Twitter are often overwhelmed with the vast amount of tweets. In this work we attempt to help users browsing a large number of accumulated posts. We propose a personalized word cloud generation as a means for users' navigation. Various user past activities such as user published tweets, retweets, and seen but not retweeted tweets are leveraged for enhanced personalization of word clouds. The best personalization results are attained with user past retweets. However, users' own past tweets are not as useful as retweets for personalization. Negative preferences derived from seen but not retweeted tweets further enhance personalized word cloud generation. The ranking combination method outperforms the preranking approach and provides a general framework for combined ranking of various user past information for enhanced word cloud generation. To better capture subtle differences of generated word clouds, we propose an evaluation of word clouds with a mean average precision measure.

Introduction

The last decade saw a social media boom when several services became widely popular and used by people all over the world. Social media services such as Facebook, Twitter, and Linkedin allow users to connect with their friends, colleagues, and other entities that are important and relevant to their interests. Active users of Twitter are often overloaded with a vast amount of posts, referred to as tweets (Bernstein et al., 2010). This overload is amplified when a user follows many users who publish excessively on a daily basis (e.g., news, politics, companies) (Douglis, 2009; Grineva & Grinev, 2012; Guo, Goh, Ilangovan, Jiao, & Yang, 2012; Hargittai, Neuman, & Curry, 2012). The work of Qu and Liu (2011) reports that an average Twitter user follows 80 users, leading to hundreds or even thousands of tweets inundating the user daily. The problem is further exacerbated if the user does not process the data in a timely manner or follows too many entities (Douglis, 2009). Thus, it is necessary to develop useful tools to help users efficiently sift through the accumulated tweets to quickly discover those that are worthy of note to the user.

To help all these users address this problem of information overload, several strategies have been proposed (Grineva & Grinev, 2012), such as adaptive facet search (Abel, Gao, Houben, & Tao, 2011) or interactive topic browsing interface that combines facet search with word cloud (Bernstein et al., 2010). The strategy of generating a word cloud to help users navigate to the most interesting tweets proposed in the work of Bernstein et al. (2010) is particularly interesting, since a visualized word cloud is very intuitive and easy for a user to employ for browsing the tweets from those followed users; Bernstein et al. (2010) showed that word clouds extended with the facet search can enhance a user experience when browsing hundreds of user tweets at once.

While much work has been done on generating tag clouds (Leginus, Dolog, & Lage, 2013; Venetis, Koutrika, &

Supporting Information
Detailed results and the source code are available at http://sourceforge.net/projects/mleginus/files/personalizedclouds/

Received December 5, 2014; revised December 9, 2014; accepted December 9, 2014

Garcia-Molina, 2011), little work has been done to address the problem of generating word clouds for Twitter users. There have been studies about exploiting word clouds generated from Twitter data as a means for navigating and summarizing a large collection of tweets for e-health surveilance purposes (Heaivilin, Gerbert, Page, & Gibbs, 2011; Lage, Dolog, & Leginus, 2014a). Further, Finin et al. (2010) exploit crowdsourcing services to perform user evaluation of word clouds generated from tweets. However, it appears that the only work directly addressing the problem of generating word clouds for users from their timeline tweets to minimize information overload is Bernstein et al. (2010). In that work, the authors generated word clouds based solely on the accumulated tweets from the followed users that a user has to process, and used statistical weighting to choose the most promising words from such tweets to form a word cloud. Such a "generic" strategy ignores much useful information about the particular user (e.g., the tweets written by the user or retweeted by the user), which intuitively contains important clues about user interests.

To address this limitation, in this paper we propose to generate a more personalized word cloud by leveraging all the personal information about a user to help generate a word cloud that can better reflect a particular user's interests and preferences. We systematically study the usefulness of three different types of past information about a user, including (a) the tweets written by the user, (b) the retweeted tweets by the user, and (c) the tweets that the user did not choose to retweet (i.e., "skipped" or "ignored" tweets, see more details in Notations), in addition to the currently accumulated tweets from those followed users that the user must process, and propose a general framework for combining all such information to generate a word cloud for the user to employ to navigate and retrieve the most relevant tweets. Using this general framework, we systematically evaluate multiple ways to leverage a user's past information to generate a personalized word cloud. We evaluate the usefulness of a word cloud by using the words in the cloud and their weights as a query to rank all the candidate tweets, and measuring the ranking accuracy by treating the retweeted tweets by the user as "relevant tweets" (i.e., the target tweets to show to the user). To the best of our knowledge, retweeted tweets are the most realistic relevance indicators when no other user relevance judgment data are available (Chen et al., 2012). The retweet action indicates that the user has carefully read the whole tweet and based on tweet relevance decided to share it with his followers. Our results show that (i) both the tweets written by the user and the retweets of the user can be exploited to improve the quality of the generated word clouds, suggesting that the proposed idea of generating personalized word clouds works very well; (ii) retweeted tweets of a user are more useful than the tweets written by the user; (iii) the skipped (ignored) tweets from the followed users are also useful for penalizing potentially distracting words when combined with the retweeted tweets; and (iv) combining all the user information performs the best.

The major contributions of this work are:

- We propose a new strategy for improving word cloud generation for Twitter users by exploiting all the past user information to generate more personalized word clouds, which outperforms the state-of-the-art method (which is nonpersonalized).
- We propose a general framework to effectively combine three types of user information, that is, tweets written by a user, tweets retweeted, and tweets ignored by the user, for generation of personalized word cloud, and systematically examine the effectiveness of each type of user information.
- We propose a new way of evaluating word clouds quantitatively based on tweet ranking, which can reveal subtle differences of different word clouds.
- We evaluate the proposed methods and come up with the following findings: (a) The retweeted tweets by users are generally more useful than the tweets written by the user for word cloud generation. (b) The ignored tweets can also be exploited to further improve the quality of word clouds. (c) A combination of all the user information yields the best results.

Related Work

User Modeling in Twitter

Similar to Bernstein et al. (2010), we consider a personalization of word clouds as an important step towards overcoming information overload and simplified retrieval of relevant tweets. Abel, Gao, Houben, & Tao (2011) explored the benefits of semantic and topic enrichment for user modeling in order to improve news recommendations on Twitter. A traditional user profile represented as a bag-of-words is extended with hashtags, detected topics, and recognized entities. The authors report a significant improvement in recommendations for user profiles that were extended with detected entities. The limitation of this approach is dependence on external named entity recognition tools. Michelson and Macskassy (2010) proposed detecting named entities for improved user modeling to minimize a vocabulary gap between user profiles and tweets for recommendations. The study leverages an external Wikipedia repository for enhanced named entity disambiguation. The method allows the extraction of high-level categories from the retrieved taxonomy. These high-level categories represent user interests. Our work differs from Abel et al. (2011) and Michelson and Macskassy (2010) in the following way: we generate model user preferences not only with the positive preferences but also read but not retweeted tweets as negative feedback, which provides more accurate modeling of user preferences. These ignored tweets are presented to the user before or after the retweeted tweets in the user timeline, which ensures that the user has skimmed through them but found them irrelevant.

Another direction for personalized information retrieval is the exploitation of other user preferences—an application of the collaborative filtering method to better recommend relevant tweets for end users. Chen et al. (2012) proposed a collaborative recommendation of tweets that is motivated by

DOI: 10.1002/asi

the assumption that those users who retweeted similar tweets in the past are very likely to retweet similar statuses in the future. Yu, Shen, and Xie (2013) exploit other users' preferences to recommend relevant tweets for a given user. The most similar users are selected according to the calculated latent topic modeling similarity and connection-based similarity with respect to a given user. That direction is not the focus of this work, but the proposed content-based personalization methods could be combined into hybrid personalization systems incorporating content-based preferences of a specific user and collaborative interests of similar users. Lage, Durao, and Dolog (2012) explore group-based, multicriteria recommendation on microbloging services where individual user preferences are extracted from user activities and linked content is aggregated together with the preferences of other users in the group of followers exploiting different group preference aggregation strategies. In order to sufficiently model group preferences when only a limited number of group members' past actions is available, the proposed negative implicit feedback-based methods from this study could be applied to enhance the quality of recommendations.

We utilize Twitter users' past actions to enhance the quality of generated word clouds. Most of the content-based recommendation systems (Abel et al., 2011; Chen, Nairn, Nelson, Bernstein, & Chi, 2010) utilize the content of user retweets and user self-created tweets. Indeed, these user actions are useful for improving the quality of personalization algorithms for the Twitter platform. In our work, we analyze which of these past user actions, that is, user retweets and user self-created tweets, is the most useful source of user preferences for enhanced word cloud generation, which is a different goal from what has been studied in existing works. We assume that the user creates and retweets content that is relevant with respect to his interests and preferences. In other words, we consider these user actions as positive relevance feedback on Twitter messages. We also address the novel question of whether it is possible to derive a user negative relevance feedback to enhance the personalization process. Many studies from information retrieval research have focused on utilizing implicit user feedback in order to enhance the quality of a web search. The advantage of implicit feedback is that users do not have to make any additional effort (Shen, Tan, & Zhai, 2005) to indicate a document's relevance. Usually, information retrieval systems exploit user clickthrough data, previous queries, and reading information reading activities of the user. The benefit of this approach is that enhanced accuracy of information retrieval systems as implicit relevance feedback enables better understanding of user information needs (Shen et al., 2005). Our work can be regarded as a novel way of incorporating implicit feedback for word cloud generation from tweets.

Word Cloud Generation

Bernstein et al. (2010) tackle an information overload problem in Twitter. The authors propose clustering semantically similar tweets such that each aggregated cluster of tweets represents a certain topic. Topic modeling on top of individual tweets fails to provide meaningful topic terms because often terms associated with words in a tweet are recognized as topic terms. Thus, the authors exploit an external search engine for enhanced topic modeling and consequent semantic annotation of tweets. The most frequent topics are presented through facet search in combination with a word cloud interface. Participants in the user evaluation agreed about the benefits of this browsing interface. Users reported the following benefits: *quick to scan, easy to browse, enjoyable, less tedious and overwhelming than a chronological timeline*. The only reported drawback of the interface is the uncertainty of users as to whether they explored all available tweets. User preferences are derived from user past tweets and represented with term frequencies forming a bag-of-words-based profile. The relevant tweets matching the user profile are presented in the system dashboard. However, the structure of a word cloud or a facet search is not adapted with respect to the user profile. In addition, a user profile is derived only from a user's past tweets, which represent user positive preferences without consideration of user past retweets and irrelevant tweets.

Similarly, Abel, Celik, Houben, and Siehndel (2011) address the inconvenience of retrieval of a vast number of user timeline tweets. Abel et al. (2011) propose an adaptive faceted search to simplify the retrieval of user-relevant tweets. Adaptation is based on the user profile extracted from one's own published tweets and extended with time context, that is, publishing time of the user. Facet extraction is enhanced with semantic enrichment of tweets such that recognized entities from tweets and external websites linked from original tweets are exploited. Our work differs from Abel et al. (2011) and Bernstein et al. (2010), since we leverage additional user past actions as user past retweets and, in particular, user negative relevance feedback derived from seen but not retweeted tweets.

Personalized Word Clouds

The aim of this work is to analyze and identify possible approaches for enhanced personalization of word clouds given user past actions. Word cloud generation should reflect user preferences and interests such that a user can retrieve relevant tweets through a generated word cloud more easily without being overwhelmed.

Bernstein et al. (2010) tackled the problem of information overload on Twitter with the interactive topic browsing interface where the word cloud is the main component of this tool. Their approach simply presents the most popular topics from the user timeline tweets and does not consider user preferences or interests. Topics are generated from search engine results that match a given tweet. Thus, a word cloud indicates social activities of users that a given user follows. The limitation of this popularity-based word cloud generation is that users might not be able to retrieve tweets that are relevant with respect to their interests. The goal of

DOI: 10.1002/asi

this study is to analyze whether personalized word cloud generation might improve retrieval of user-relevant tweets over this standard popularity or importance-based word cloud generation. If yes, what are the best ways of leveraging user past actions to model user preferences? How to incorporate user interests into personalized word cloud generation? What is the optimal way to quantitatively evaluate the quality of word clouds? To analyze and answer these questions, let us first introduce basic notations and then define a personalized word cloud generation.

Notations

The set of tweets published by a user U is denoted as U_{tw}, the set of retweeted tweets is U_{retw}. Seen but not retweeted tweets that occur within k positions before or after a retweeted tweet from U_{retw} belong to set U_N. We assume that tweets from the set U_N are irrelevant for the given user U. All tweets from sets U_{tw}, U_{retw}, U_N are published during a period $tm_0 - tm_1$. For the sake of simplicity, a set of user past actions tweets, denoted U_P, can represent either U_{tw}, U_{retw}, or U_{tw}. The set of timeline tweets of user U is denoted as U_{tml} and these tweets are published between a period tm_1 and tm_2 where $tm_0 < tm_1 < tm_2$. All user past actions U_P were performed before tm_1, that is, from tm_0 to tm_1.

Task Definition

Given the timeline tweets U_{tml} of user U, our task is to select the most relevant top-k terms such that the selected terms are relevant and link to relevant tweets in U_{tml}, which were published between tm_1 and tm_2 by the Twitter users that user U follows. To adapt word cloud generation towards user preferences and interests, the following user past actions and corresponding tweets are exploited:

- U_{tw}—set of tweets published by user U
- U_{retw}—set of tweets retweeted by user U
- U_N—set of tweets seen but not retweeted by user U

We focus on personalization purely based on user preferences derived from user past performed actions (between tm_0 and tm_1) such as: publishing personal statuses, retweeting relevant tweets of users she/he follows, and reading information of the user. Please note that $tm_0 < tm_1 < tm_2$, hence word clouds are generated only from tweets published between tm_1 and tm_2 excluding tweets published before tm_1, that is, from which a user profile is derived.

Research Questions

To better understand and analyze the problem of personalized word cloud generation with respect to user preferences, we pose the following questions:

R1: Does personalized word cloud generation improve the retrieval of user relevant tweets?

R2: Which user preferences derived from U_{tw}, U_{retw} or combined preferences ($U_{tw+retw}$) enhance the quality of the word cloud the most?

R3: Does implicit relevant feedback derived from seen but not retweeted tweets U_N enhance personalized word cloud generation? If yes, what is the best way to incorporate these user past actions?

Graph-Based Word Cloud Generation

To incorporate user preferences derived from past user actions for adapted word cloud generation, we utilize graph-based ranking. These methods are capable of incorporating user preferences and consequently bias word cloud generation. In our previous work (Leginus et al., 2013), we showed the benefits of leveraging graph-based methods for more relevant tag cloud generation from folksonomy data. Similarly, several studies prove the advantages of using graph-based methods for keyword or sentence extraction from text data (Bellaachia & Al-Dhelaan, 2012; Inouye & Kalita, 2011; Mihalcea & Tarau, 2004; Wu, Zhang, & Ostendorf, 2010). The advantage of this work is the possibility of incorporating the findings into any graph random-walk based summarization method for Twitter data.

Graph creation. In order to perform a graph-based ranking of words, we extract terms from the underlying tweets. We consider each term as a graph vertex. If two vertices co-occur at least α times, we consider these two terms similar. Eventually, for each similar word pair, two directed edges are generated, $t_1 \rightarrow t_2$ and $t_2 \rightarrow t_1$.

Graph-based ranking. To rank the relevance of terms with respect to user preferences, we utilize graph-based ranking that simulates a stochastic process, that is, random traversal of the terms in the graph. In this work we exploit the topic-sensitive PageRank algorithm (Haveliwala, 2003), but any other algorithm based on random traversal of the graph such as HITS,k-step Markov Chain and others can be employed (Leginus et al., 2013; White & Smyth, 2003).

The aim is to estimate the importance of a term t with respect to user preferences. To bias ranking towards user preferences, a vector of prior probabilities $\mathbf{p_{u_p}}$ is defined. For not personalized graph-based ranking, we set each entry in $\mathbf{p_{u_p}} = \{p_1 \dots p_{|V|}\}$ to $\frac{1}{|V|}$ where V is the set of all graph vertices. The sum of prior probabilities in $\mathbf{p_{u_p}}$ equals 1. A random restart of stochastic traversal of the graph is assured with a back probability β, which determines how often a random traversal restarts and jumps back to user-preferred terms that represent user preferences. Therefore, the β parameter allows one to adjust the bias towards user preferences or towards the vertices that are globally relevant in the underlying graph.

To simulate random traversal of the graph, iterative stationary probability is defined as:

$$\pi(v)^{(i+1)} = (1-\beta)\left(\sum_{u=1}^{d_{in}(v)} p(v \mid u)\pi^{(i)}(u)\right) + \beta \mathbf{p_{u_p}}$$

DOI: 10.1002/asi

where $\pi(v)^{(i+1)}$ is a probability of visiting node v at time $i + 1$, $d_{in}(v)$ is set of all incoming edges to node v, and $p(v|u)$ is a transition probability of jumping from node u to node v. In this work a transition probability is set to $p(v \mid u) = \frac{1}{d_{out(u)}}$ for nodes v that have an ingoing edge from node u, otherwise $p(v|u)$ equals 0.

The resulting rank of term t biased towards user preferences after convergence is considered as relevance of t, that is;

$$I(t \mid u_p) = \pi(t)$$

Encoding user information. We incorporate user preferences and interests to adapt word cloud generation through setting prior probabilities vector $\mathbf{p_{u_p}}$. Based on the given source of user past actions, we analyze the content of past user published tweets U_{tw}, past retweets U_{retw}, and past "ignored" tweets U_N to derive user preferences. In order to not propagate common and not very discriminative words, we compute the term frequency-inverse document frequency (tf-idf) score for each term occurring within the tweets of the user past actions. Due to a short length of tweets, we exploit a hybrid TF-IDF scoring as introduced in Inouye and Kalita (2011).

$$TF - IDF_{t_i} = tf_{t_i,U_P} * \log_2 \frac{N}{df_i} \quad (1)$$

where tf_{t_i,U_P} represents term frequency of a term t_i within user past actions tweets U_P and N represents the number of all tweets from user past actions tweets, and df_{t_i} is the number of tweets from user past actions tweets that contain term t_i. This scoring computes term frequency within all tweets of user past actions U_P, which are considerd one document. Document frequency is counted in the standard way such that each tweet is considered as a distinct document.

Calculated tf-idf scores are then leveraged for prior vector $\mathbf{p_{u_p}}$ creation such that each term that has occurred in user past action tweets U_P and occurs in the tweets from timeline U_{tml} is represented with the normalized computed tf-idf score. The original tf-idf scores are normalized with the sum of all tf-idf scores such that prior vector entries sum to 1. When a term does not occur within the user past action tweets U_P but occurs only in tweets from the user timeline U_{tml}, we set a corresponding entry in $\mathbf{p_{u_p}}$ to 0.

The advantage of graph-based ranking methods is that even when user preferences are not sufficiently defined, that is, a cold start problem, or the tweets from the user timeline U_{tml} are semantically different from the user past action tweets U_P, the method will promote globally important terms from the user timeline U_{tml}, that is, user U will be able to explore globally important terms from the underlying tweets. The random traversal of the graph ensures that from a few terms that represent user preferences other globally important terms, nodes, can be traversed and consequently highly ranked.

Convergence

The described graph creation ensures that no dangling nodes are introduced into the graph. The prior vector $\mathbf{p_{u_p}}$ in both cases (personalized and not personalized forms) consists of nonnegative elements whose L1 norm is 1; therefore, a graph-based ranking always converges to the same unique stationary distribution (Haveliwala & Kamvar, 2003). Thus, whenever we generate a word cloud with a certain prior vector on top of the same set of tweets, the word cloud will be always the same for the given prior vector.

Implicit Feedback

To maximize benefits of available user past actions, we propose a general framework which combines user positive preferences derived either from user past tweets U_{tw} or from user past retweets U_{retw} with user negative preferences U_N and in such a way that enhances the personalization of word clouds. We design two different methods for incorporating negative user preferences into graph-based word cloud generation methods. The first is based on modyfing prior probabilities vector $\mathbf{p_{u_p}}$ with the approach similar to the Rocchio model. The second method performs reranking of the top-k terms to penalize irrelevant terms from read but not-retweeted tweets U_N.

Preranking combination (PRC). Prior probability vector $\mathbf{p_{u_p}}$ in PageRank enables the bias ranking of terms towards user preferences. To incorporate user negative preferences, we propose to leverage the Rocchio algorithm. The motivation is to bias ranking of words towards salient words from user relevant tweets U_P which are not occurring within irrelevant tweets U_N. We calculate prior probability $w(t_i)$ for each term $t_i \in W$ in the following way:

$$w(t_i) = \begin{cases} \frac{s(t_i)}{\sum_{t_j \in W} s(t_j)} & s(t_i) > 0 \\ 0 & s(t_i) < 0 \end{cases}$$

where W is a set of all terms extracted from U_{tml}. Scoring function $s(t_i)$ is defined as:

$$s(t_i) = \left(\frac{1}{|U_P|} \sum_{tw_j \in U_P} tf(t_i, tw_j) \cdot \lambda_{tw_j} - \frac{\gamma}{|U_N|} \sum_{tw_k \in U_N} tf(t_i, tw_k) \cdot \mu_{tw_k} \right) \cdot idf(t_i) \quad (2)$$

where $tf\ (t_i,\ tw_j)$ is term frequency of t_i in tweet tw_j. $s(t_i)$ can then be normalized into probabilities, which will be used to set $\mathbf{p_{u_p}}$, thus injecting the "feedback bias" into PageRank.

Parameter γ determines the extent to which the terms from irrelevant tweets are penalized. The higher value causes terms occurring in both sets to be suppressed and

DOI: 10.1002/asi

only words occurring within relevant tweets are amplified and propagated to the graph-based word cloud generation. Weights μ_{tw_k}, λ_{tw_j} allow promoting certain tweets in either relevant tweets or irrelevant tweets sets and consequently influence word cloud generation.

Possible usage scenarios of weighting terms frequencies with weights μ_{tw_k}, λ_{tw_j} could be:

- Confidence about seen but ignored tweets—the more certain it is that a given tweet has been seen but ignored, the more this information should be exploited. Conversely, when the system is not certain whether a given tweet has been read the weight might be smaller. Confidence whether a user has read a certain tweet could be defined as a ratio of passed time before or after a specific user retweet action. Alternatively, a rank distance (tweets have to be presented in chronological order) between a retweeted tweet and a certain tweet could capture the confidence of reading a given tweet.
- Time—if a certain user tends to change user preferences and interests often, then more recent relevant tweets should be weighted higher.
- Similarity—when user past retweets or one's own published tweets are available it is possible to calculate a similarity between considered negative tweets and user past relevant tweets. Tweets more similar to user past tweets are more likely to be relevant and thus they should be weighted less in penalizing part of the algorithm.
- Relevance feedback—generate personalized word clouds with respect to user preferences, retrieve relevant tweets that match a word cloud query. Top k retrieved tweets can be used for measurement of a tweet's similarity with irrelevant past tweets or relevant past tweets.

Ranking combination (RC). Once graph-based ranking is performed, we penalize important terms from the irrelevant tweets U_N. We class irrelevant tweets as those presented in the user U timeline above or below a retweet action of a user U. We compute graph-based ranking with negative preferences $\mathbf{p}_{\mathbf{negPref}}$. The negative prior probabilities vector $\mathbf{p}_{\mathbf{negPref}}$ is derived from U_N tweets and it is encoded similarly as detailed in the earlier subsection, Encoding user information. The final ranking score for each term $t_i \in W$ is computed in the following way:

$$finalScore(t_i) = \alpha * I(t_i \mid \mathbf{p}_{\mathbf{pref}}) - (1-\alpha) * I(t_i \mid \mathbf{p}_{\mathbf{negPref}})$$

where $I(t_i|\mathbf{p}_{\mathbf{pref}})$ is a PageRank score of term t_i biased towards $\mathbf{p}_{\mathbf{pref}}$ and similarly $I(t_i|\mathbf{p}_{\mathbf{negPref}})$ a PageRank score biased with respect to the negative preferences calculated from all the tweets that were read but not retweeted by the user U_N.

The aim of this method is to prefer terms relevant with respect to user preferences and at the same time terms should not be relevant within tweets read but not retweeted by the user.

Possible Applications

The proposed approach of word cloud personalization has multiple applications. The first is a visualization of user relevant keywords from the underlying tweets matching user preferences. The personalized word cloud enables a user to browse tweets, obtain an understanding of underlying tweets collection, as well as discover new interesting and relevant keywords that might be further exploited for keyword searches. Personalized word clouds are mainly intended for user homeline tweets browsing (Bernstein et al., 2010). Second, the approach of word cloud personalization could be easily applied when browsing and exploring a large collection of tweets (Lage et al., 2014a; O'Connor, Krieger, & Ahn, 2010).

Experimental Results

Evaluation Methods

We evaluate the proposed methods with the same Twitter data set as in Lage, Dolog, and Leginus (2014b), which was harvested during a period from September 17 to October 16, 2011. We reconstruct for each user U a timeline consisting of the tweets from users he/she follows. Tweets that were retweeted by a user U are considered relevant ones and thus utilized for the evaluation of word cloud generation.

This Boolean relevance indicator provides a rough insight about the user's interest about the retweeted tweet. However, the lack of retweets does not necessarily reflect that tweets are not relevant for the user.

This is a known challenge when evaluating personalized Twitter services and no additional user relevance judgment data are available (Chen et al., 2012).

We designed the evaluation procedure as follows:

1. Group the tweets from a user *U's* home timeline, that is, tweets published by the users she/he follows from a predefined time window. In this evaluation, we aggregated published tweets according to the week in which they were published. Think of this as a user U logging into Twitter once a week.
2. For each user *U's* week timeline tweets, we generated a personalized word cloud with 10, 15, or 20 terms. User preferences are derived from the user's past actions (set of tweets published by user U, tweets retweeted by user U, and tweets seen but not retweeted by user U) that were performed before the given week. Thus, **word clouds are generated from user *U's* week timeline tweets excluding any user past actions related tweets**. User *U's* retweets for the given week are treated as test data.
3. For each personalized word cloud we measure mean average precision and relevance where relevant tweets are those that were retweeted by user U during the given week.

We filter user week timelines such that each user timeline for a given week contains more than 1,000 tweets from other users to sufficiently simulate information overload (please see the discussion about information overload impact on the word cloud personalization in Levels of Information Overload). We only consider user *U's* week timelines with at least 20 retweets of user U to measure an effect of personalization methods. At the same time, a user U has to have at

DOI: 10.1002/asi

least 20 retweets before a given week to derive user preferences. The final evaluation set contains 143 different week timelines for 107 distinct users.

The evaluation framework was implemented in Java and the source code can be downloaded from the main author's personal website.[1]

Evaluation Metrics

In this study, we measure the relevance of generated word clouds as done in the existing work (Leginus et al., 2013; Venetis et al., 2011). Further, we propose a new way of measuring the quality of word cloud personalization through the mean average precision metric. The metric provides a more thorough evaluation of word clouds from a user's perspective as well as being sensitive to the weights of words.

Relevance is defined as:

$$Relevance(WC_k) = avg_{t_i \in WC_k} TermRelevance(t_i, U_p)$$

where relevance for each term in the cloud is defined as:

$$TermRelevance(t, U_p) = \frac{|T_t \cap T_{retw}|}{|T_t|}$$

and T_t is a set of tweets containing term t and T_{retw} is a set of retweeted tweets of user U during a certain time period (for this work, a 1-week period is used).

However, this measure does not capture ordering differences of words within the cloud and considers each term as a single query. The assumption that terms depicted in the cloud are of equal importance is often invalid. We believe that word weights and their order are important aspects of word clouds where better ranked terms might be more visible, that is, larger font size or better position. To address this, we propose a methodology for the evaluation of personalized word clouds.

We consider a generated word cloud as a query which should retrieve user-relevant tweets, that is, user retweets. Therefore, a more personalized word cloud should link to more relevant user tweets. To model and measure this, we propose the following:

1. For given terms and corresponding weights of a word cloud WC_k, create a query vector Q_{WC_k} with normalized weights. Each entry of the query vector Q_{WC_k} represents the importance of a term from the word cloud WC_k with the normalized weight, that is, more important terms from the word cloud are represented with higher weights.
2. Rank and retrieve top-k tweets matching a given query Q_{WC_k}.
3. Measure mean average precision where each retweeted tweet is considered relevant.

Ranking of relevant tweets with respect to a given query Q_{WC_k} is computed with the standard information retrieval function OKAPI BM25, which can be defined as:

$$S(tw, Q_{WC_k}) = \sum_{q_i \in Q_{WC_k} \cap tw} c(q_i, Q_{WC_k}) \cdot TF(q_i, tw) \cdot IDF(q_i) \quad (3)$$

where

$$TF(q_i, tw) = \frac{f(q_i, tw) \cdot (k_1 + 1)}{f(q_i, tw) + k_1 \cdot \left(1 - b + b \cdot \frac{|tw|}{avgtwl}\right)}$$

$$IDF(q_i) = \log \frac{N - n(q_i) + 0.5}{n(q_i) + 0.5}$$

and $f(q_i, tw)$ is a q_i term frequency within a tweet tw, $|tw|$ is the length of a given tweet tw, $avgtwl$ is average length of tweet within the corpus, N is a total number of tweets in the corpus, and $n(q_i)$ is the number of tweets that contain the term q_i. To capture the importance of a word from the generated word cloud, we multiply the whole relevance score for a given term with the word cloud weight $c(q_i, Q_{WC_k})$ for the given term q_i. The function $c(q_i, Q_{WC_k})$ returns a weight of the term q_i from the query vector Q_{WC_k}, which corresponds to the term weight from the word cloud WC_k. We set the following values for parameters $k_1 = 1.2$ and $b = 0.75$.

We measured the average precision at K for the retrieved top K list of ranked tweets with respect to the given word cloud. Consequently, we measured the mean average precision for all generated word clouds. The average precision of top K ranked tweets with respect to the word cloud is calculated as follows:

$$AP@K(Q_{WC_k}) = \frac{\sum_k^K (P(k) \cdot rel(k))}{\#relevanttweets}$$

where $P(k)$ is the precision at k-th position in the ranked top K list and $rel(k)$ is 1 if the tweet at rank k is relevant, that is, retweeted, otherwise $rel(k)$ is 0 and $\#relevanttweets$ is the number of retweeted tweets within the top K list. Mean average precision (MAP) is defined as:

$$MAP@K = \frac{\sum_{Q_{WC_k}} \in AWC_k AP@KQ_{WC_k}}{|AWC_k|}$$

where AWC_k is the set of all generated word clouds and $AP@KQ_{WC_k}$ is average precision for the given word cloud Q_{WC_k}. In this work, we measure MAP at 20 under the assumption that it represents a reasonable cutoff for the number of tweets that a user would likely retweet.

A significant advantage of this new measure over the existing measure is that it is not only sensitive to terms that are included in the word cloud, but also sensitive to the weights of terms from the word cloud as this affects the relevance scores of retrieved tweets (as shown in Formula [3]). Therefore, a better word cloud is one where ranked terms link to many user-relevant tweets.

Despite the aforementioned relevance limitations, we also report relevance results of generated word clouds for the following reasons. Leginus et al. (2013) position

[1]http://sourceforge.net/projects/mleginus/files/personalizedclouds/

DOI: 10.1002/asi

relevance as a meaningful measure of a tag cloud quality and present relevance benefits over other tag cloud metrics such as coverage or overlap. Further, we are interested in understanding whether the proposed measurement of mean average precision correlates with relevance measure.

Tweets Preprocessing

Each tweet is preprocessed. Special nonalphabetical characters expect that #, urls, user mentions @*user*, and stop words are removed. The size of stop words is 6,052 words and it includes different acronyms often used on social networks.[2] We do not generate personalized word clouds for those users where a ratio of non-English tweets during a considered week was more than 10%.

Baseline Selection Techniques

We compare the performance of personalized word cloud generation methods with the following algorithms proposed in existing work:

Term frequency—inverse document frequency selection (TF-IDF). The method proposed by Inouye and Kalita (2011) is an adapted version of term frequency—inverse document frequency weghting. Due to the short length of tweets, standard term frequency for most of the terms is 1. The authors suggest redefining a notion of document such that more tweets are considered as one document for term frequency calculations but inverse document frequency is calculated the standard way, such that each tweet is treated as a single document. This approach is able to differentiate term frequencies in comparison with the standard term frequency—inverse document frequency weighting. In this evaluation, TF-IDF ranks each term t_i from the timeline tweets U_{tml} according to Formula (1). All tweets from the U_{tml} are considered a single document when calculating term frequencies. These values are sorted in descending order. The top-k terms with the highest scores are selected for the final word cloud.

Not-personalized PageRank (NoPerPR). This method was originally proposed in Leginus et al. (2013) to estimate tags relevance with respect to a certain query and it outperformed in terms of relevance several tag selection approaches. The method estimates global terms importance within the graph created from user timeline tweets without personalization.

Parameters setting of graph-based techniques. We generate word cloud graphs from the extracted terms of user timeline tweets U_{tml} (similar to our previous work, Leginus et al., 2013). Due to low co-occurrence of terms, we set threshold α to 0. Thus, for each two terms that co-occur within user timeline tweets U_{tml}, the method generates two directed edges $t_1 \rightarrow t_2$ and $t_2 \rightarrow t_1$.

We set the following parameters for used graph-based methods:

- Prior probabilities for NoPerPR are defined as $\frac{1}{|V|}$ for each word $t_i \in W$.
- Prior probabilities for PerPR are defined according to 3.4.
- A back probability β for *NoPerPR* and all variations of *PerPR* algorithms is set to $\beta = 0.85$. *P*arameter for PRC is $\gamma = 0.75$. For penalized postprocessing, we set $\alpha = 0.8$. The impact of parameter settings on the word cloud generation will be discussed in Implicit Feedback and Parameter Setting.

Personalized Word Clouds

The comparison of baseline techniques and personalized graph-based methods is presented in Table 1. Baseline methods TF-IDF and NoPerPR attain low mean average precision as well as relevance. NoPerPR slightly outperforms TF-IDF, which corresponds with the findings from Wu et al. (2010). When comparing baseline methods with the two different variations of PerPR, it is obvious that leveraging user past actions improves word cloud generation. The best personalized method attains almost five times higher mean average precision and similar results are attained when measuring relevance of generated word clouds. Hence, this empirical result clearly answers question *R*1 that personalized word clouds improve precision and relevance of generated word clouds, which implies improved retrieval of user relevant tweets from user timeline tweets. Please note that the intention of this work is to find out which ways are the most efficient when personalizing word clouds. The initial comparison with nonpersonalized state-of-the-art approaches is meant to illustrate an importance of personalization for users who are overloaded with many tweets in their timelines.

Further, we analyze the impact of distinct user past actions on the word cloud generation. The results prove that both the user's past tweets (written by the user) and the user's past retweets are useful for personalized word cloud generation. We found that a user's past retweets better characterize user preferences and interests. The explanation for this is that users often publish tweets that are not indicative

TABLE 1. Mean average precision and relevance of TF-IDF, nonpersonalized Pagerank (best baseline method), and two variations of personalized PageRank algorithm leveraging different user past actions.

Method	Mean average precision			Relevance		
	#10	#15	#20	#10	#15	#20
TF-IDF	0.062	0.057	0.040	0.026	0.028	0.028
NoPerPR	0.068	0.062	0.063	0.029	0.029	0.029
PerPR (U_{tw})	0.162	0.181	0.198	0.067	0.0628	0.06
PerPR (U_{retw})	**0.350**	**0.344**	**0.346**	**0.154**	**0.136**	**0.127**

[2]http://www.noslang.com/search.php

DOI: 10.1002/asi

of their preferences. Naaman, Boase, and Lai (2010) report that users often tweet about the following topics: *Me Now (≈40% of all tweets)*, for example, tired and upset, *Presence maintenance (≈5%)* (e.g., i am backk), or *Statements and random thoughts (≈25%)* (e.g., the sky is blue in the winter here) which are clearly not very informative about user preferences and interests. The naive approach of combining user past published tweets and past retweets does not outperform personalization based on user past retweets. Thus, the answer to the posed question *R*2 is that user past retweets are more useful than user past tweets. The naive union of user past tweets and user past retweets does not outperform $PerPR(U_{retw})$.

Implicit Feedback and Parameter Settings

We showed that a user's past retweets enhance personalized word cloud generation in comparison to the user past published tweets. When a user retweets a certain tweet, he has to first read it. The current user interface of Twitter presents tweets from the user timeline in chronological order. This allows us to assume that tweets presented before or after a retweeted tweet had to be read by the user. Thus, we assume that those *read but not retweeted* tweets are not sufficiently relevant. We utilize this information to derive user negative preferences and, consequently, improve personalized word cloud generation. In the following paragraphs we present the results obtained using such a negative feedback strategy.

The results are presented in Table 2. Both proposed methods for integration of negative feedback preranking combination (PRE) and postranking combination (RC) methods outperform the best PerPR (U_{retw}) method, which is based only on positive user past actions, that is, user's past retweets. The best results are attained by the RC method, which outperforms PerPR (U_{retw}) with improvements of 8.29% for 10 terms, 6.98% for 15, and 7.8% for 20 terms in the cloud. The RC method outperforms PRE ranking by 4.41% for 10 terms, 0.82% for 15, and 3.9% for 20 terms in the cloud. This suggests that combining the rankings generated by graph-based ranking is generally more effective than combining the term weights first and then feeding the term weights into the graph-based ranking algorithm (i.e., postranking combination works better than preranking combination). One possible explanation of this is that the graph-based ranking algorithm may have a "smoothing" effect on the user's feedback information and, thus, if the negative feedback information is not reliable, it would not directly affect the final ranking in the case of the postranking combination as in the preranking combination.

Both methods are parameter-dependent and therefore we also analyze the impact of these parameters on the quality of word clouds. First, we present how a skipped tweet set parameter (the number of tweets presented before and after a given retweet) affects the overall mean average precision. The parameter conditions of how many skipped tweets before and after a given retweet are used as a source of user negative preferences for both methods. For instance, if the parameter is set to 40, we then select 40 tweets that occurred before and after each user retweet. Union of the obtained skipped tweets is then input to the proposed methods. This selection is based on the assumption that tweets are presented to the end user in chronological order as in the Twitter interface.

Figure 1 presents the relation between the parameter and mean average precision when the generated word cloud contains 10 terms. Both methods improve mean average precision of generated word clouds as the number of considered tweets increases. The best performance is attained with 40 tweets before and 40 tweets after a given retweet for the generation of negative user preferences. As the parameter size grows, the mean average precision slightly decreases. However, the decreased mean average precision is still slightly better than $PerPR(U_{retw})$. These results strongly suggest that the negative feedback information is useful.

Further, both proposed methods have parameters to control the extent to which the derived negative information should be leveraged (i.e., γ and α). For the PRE method, we found the best performance with $\gamma = 0.75$. For lower values the performance decreased and with parameter values lower than 0.25 the quality of word clouds is worse than

TABLE 2. Mean average precision and relevance improvements with negative feedback (RC, ranking combination, PRC, preranking combination).

Method	Mean average precision			Relevance		
	#10	#15	#20	#10	#15	#20
PerPR (U_{retw})	0.350	0.344	0.346	0.154	0.136	0.127
RC (U_{retw})	**0.379**	**0.368**	**0.373**	**0.186**	**0.164**	**0.149**
PRC (U_{retw})	0.363	0.365	0.359	0.174	0.157	0.143

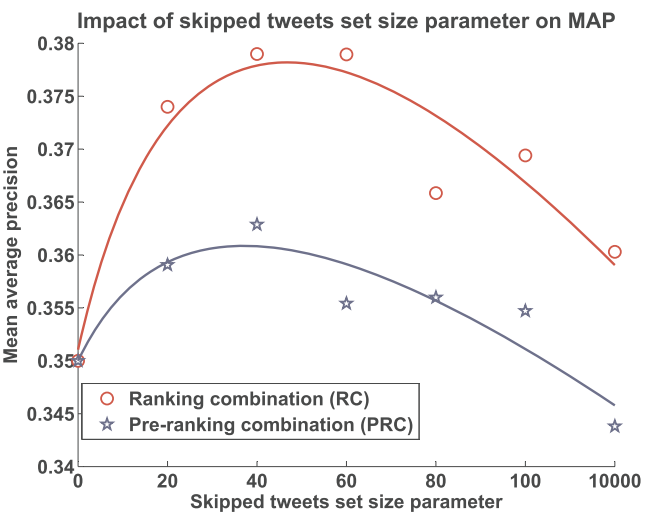

FIG. 1. Impact of skipped tweet set parameter on mean average precision. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

DOI: 10.1002/asi

FIG. 2. Word cloud examples demonstrating improvements in personalized methods.

$PerPR(U_{retw})$. For the RC method, when α is set to smaller values, penalization of terms biased towards derived negative preferences is more aggressive. The best results were attained with $\alpha = 0.8$. When decreasing the value of parameter α the performance drop is smaller and still outperforms $PerPR(U_{retw})$. Once again, we see strong support for the usefulness of negative feedback.

To better illustrate the positive impact of the personalized word cloud generation, we present four different word clouds generated for the same user on top of the same week user timeline tweets (week 40 in 2011). Each word cloud is produced by different personalization methods except the first one. The first word cloud depicted in Figure 2 is generated by the NoPerPR method and it reflects the most important events of week 40 in 2011, such as the death of Steve Jobs from Apple and protests organized by the Occupy Wall Street movement. The second word cloud is personalized leveraging user past tweets. The user was involved in subjective discussions with other Twitter users about different news, conflicts, and events in the world. Thus, the generated word cloud contains many subjective terms such as *good, bad, truth, question*. Further, the user expressed his/her opinion (before week 40) that the conflict between Israel and Palestine is similar to the problem of apartheid in South Africa. Therefore, terms extracted from tweets about the Israel and Palestine conflict and South African apartheid are depicted in the cloud. The third cloud is personalized with respect to the user's past retweets and it reflects user interests in the conflict between Israel and Palestine as well as curiosity about the Occupy Wall Street movement and their protests. The word cloud still contains some general terms like *news, people, police* which do not have high discriminative value. The fourth word cloud generated by the RC method combines user past retweets for personalization together with the negative implicit feedback from skipped tweets. Obviously, the approach helps to improve the quality of the word cloud by replacing nondiscriminative terms with terms related to the Wall Street movement protests such as *#usdor, occupy*, and *violence*.

The findings show that a combination of positive feedback signals (retweeted tweets) and negative feedback signals (possibly viewed but not retweeted tweets) is the most effective strategy for feedback. We now examine whether we can further improve performance by combining such a comprehensive feedback strategy (positive plus negative) with the use of the tweets the user has written in the past. To achieve this, we leverage the RC method and extend it to the following form:

$$finalScore(t_i) = \alpha * I(t_i \mid \mathbf{p_{pref}}) - I(t_i \mid \mathbf{p_{negPref}}) + \eta * I(t_i \mid \mathbf{p_{twPref}}) \quad (4)$$

where $\mathbf{p_{twPref}}$ represents a prior vector generated from user past tweets, performed in the same way as described in the Encoding User Information (section). The proposed method slightly outperforms the best-performing method RC (U_{retw}) for word clouds with 15 and 20 terms (see Table 3). The parameter α is set to 0.8 and η to 0.3. This suggests that combining *all* the user information, including both retweeting behavior and the tweets written by the user, works the best.

TABLE 3. Mean average precision improvements with RC method that combines user past tweets, retweets, and skipped tweets altogether.

Method	Mean average precision		
	#10	#15	#20
RC (U_{retw})	**0.379**	0.368	0.373
RC ($U_{retw+tw}$)	0.363	**0.373**	**0.380**

Levels of Information Overload

To analyze how a number of tweets presented to a user in the timeline influences word cloud generation, we performed a limited evaluation. We iteratively changed a number of tweets presented in the timeline for a user per week, that is, selecting different user week timelines with approximately similar number of tweets as the defined threshold (10 distinct timelines for each iteration). The threshold defining a minimal number of tweets parameter has been iteratively changed ranging between 100 to 1,000 with steps of 100. It is observed that word clouds generated from 100 or 200 tweets timelines attain high levels of MAP for nonpersonalized word cloud generation (0.76 for 100 tweets timelines and 0.38 for 200 tweets timelines). However, the personalized algorithm (RC method) outperformed the nonpersonalized alternative, the relative improvements were 16% and 34%, respectively, for 100 and 200 tweets timelines. The differences between standard and personalized algorithms become much more obvious for word clouds generated from 300 and more tweets timelines, that is, the relative improvements of personalized word cloud generation are almost threefold. When word clouds are generated from more than 500 tweets a significant drop of accuracy is observed for nonpersonalized version (MAP levels below 0.15). Therefore, a personalized approach of word cloud generation could be exploited when a user is exploring and browsing more than 500 tweets at once, for example, scenarios when a user logged in after a certain time and more than 500 tweets were not yet seen by the user.

DOI: 10.1002/asi

Discussion

Several studies about personalized Twitter services exploit "user retweet action" as a relevance indicator for evaluation purposes (Chen et al., 2012; Uysal & Croft, 2011; Yan, Lapata, & Li, 2012). The assumption that retweeted tweets are interesting and relevant for the user has been confirmed with small user studies (Uysal & Croft, 2011; Yan et al., 2012). Conversely, Rout, Bontcheva, and Hepple (2013) point out that a user's retweet activity is not always a reliable indicator of relevance. According to the performed user study, only 66% of interesting tweets classified by evaluation participants were retweeted by anyone at all. Frequently, Twitter users do not retweet interesting tweets that are part of a personal communication thread, as it does not make sense to broadcast them to users outside of the conversation scope. Thus, our evaluation results may underestimate the actual utility of the generated word cloud since some assumed nonrelevant tweets might actually be interesting to the user. A user evaluation should be performed to further measure whether the personalized word cloud generation enhances information access and limits information overload.

The measured levels of mean average precision are lower than in standard information retrieval evaluations. This lower accuracy might be caused by many meaningless tweets matching user preferences. According to Analytics (2009), 41% of tweets were categorized as pointless babble, 38% as conversational, 9% as pass along value, etc. Hence, only a small number of tweets is considered relevant, which naturally leads to lower levels of accuracy. This work attains similar and greater accuracy than reported in the work of Soboroff, Ounis, Lin, and Soboroff (2012), where several state-of-the-art ranking and filtering algorithms were exploited. When exploiting personalized word cloud generation, one should consider a filter bubble problem (Nguyen, Hui, Harper, Terveen, & Konstan, 2014), that is, a state when a user is exposed only to limited fragments of content which prevents an exploration of content not matching user views. A graph-based nature of personalized word cloud generation provides a means to minimize the effect of filter bubble. First, a parameter β can be tuned to define the extent of personalization, that is, the parameter defines whether to prefer user-related topics or globally important topics. Second, several diversification algorithms (e.g., DivRank; Mei, Guo, & Radev, 2010) were proposed to further diversify graph-based ranking that could be applied for diversification of personalized word cloud generation.

Future Work

In future work, we will focus on how to explore different weighting techniques that will allow better leverage of user past actions, for example, certain non-retweeted tweets might be more relevant than others. Further, we would like to extend the approach with nonlexical information to better capture and propagate user preferences when generating word clouds. We envision the following possible extensions with different nonlexical information:

- **Hashtags** extension enlarges an original set of related relevant tweets, that is, retweeted tweets. Hashtags mentioned in retweeted tweets are exploited for retrieval of additional tweets to enrich a user profile. The intuition is that tweets containing a hashtag that was mentioned in a user retweet should be related and relevant for the user. We found that with this naive extension we can slightly improve MAP. For instance, when extending RC (U_{retw}) with hashtags we attained the following relative improvements: 3.2% for word clouds with 10 terms, 8.8% for 15 terms, and decreased performance for the clouds with 20 terms. Limitation of hashtags extension is a sparse mention of hashtags within tweets as well as many hashtags that are widely popular and not very informative, for example, *#news, #breaking*. Hence, more sophisticated methods should be developed to detect only specific and relevant hashtags and then exploit them for user profile enrichment.
- **Recognized named entities** extension is the same as **Hashtags extension** except named recognized entities are used instead of hashtags.
- **Favorite followees** extension filters favorite followees and promotes salient terms from their tweets into a user profile.

Conclusion

We propose and evaluate a graph-based approach for enhanced generation of personalized word clouds for Twitter users. The proposed method outperforms baseline methods. We propose a framework to effectively combine user past actions such as user past tweets, user past retweets, and skipped tweets by the user for improved word cloud generation. In addition, we propose a new way to measure the quality of word clouds that better reflects differences between generated word clouds than relevance measure. The main findings of this study are: (a) Personalized word cloud generation improves accuracy and relevance and consequently enables easier retrieval of user relevant tweets; (b) User past retweets are more useful for a personalized generation of word clouds than tweets published by the user; and (c) Non-retweeted tweets when properly combined with user past retweets further enhance personalized word cloud generation.

References

Abel, F., Celik, I., Houben, G.-J., & Siehndel, P. (2011). Leveraging the semantics of tweets for adaptive faceted search on Twitter. In L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N. Noy, & E. Blomqvist (Eds.), The semantic web–ISWC 2011 (pp. 1–17). Berlin Heidelberg: Springer.

Abel, F., Gao, Q., Houben, G.-J., & Tao, K. (2011). Analyzing user modeling on Twitter for personalized news recommendations. In J. Konstan, R. Conejo, J. Marzo, & N. Oliver (Eds.), User modeling, adaption and personalization (pp. 1–12). Berlin Heidelberg: Springer.

Analytics, P. (2009). Twitter study–august 2009. San Antonio, TX: Pear Analytics. Retrieved from www.pearanalytics.com/blog/wp-content/uploads/2010/05/Twitter-Study-August-2009.pdf

Bellaachia, A., & Al-Dhelaan, M. (2012). Ne-rank: A novel graph-based keyphrase extraction in Twitter. In N. Zhong, Z. Gong, Y.-M. Cheung, P. Lingras, P.S. Szczepaniak, & E. Suzuki (Eds), Proceedings of the the 2012 IEEE/WIC/ACM international joint conferences on web intelligence and intelligent agent technology (Vol. 01, pp. 372–379). Washington, DC, USA: IEEE Computer Society.

DOI: 10.1002/asi

Bernstein, M.S., Suh, B., Hong, L., Chen, J., Kairam, S., & Chi, E.H. (2010). Eddi: interactive topic-based browsing of social status streams. In K. Perlin, M. Czerwinski, & R. Miller (Eds), Proceedings of the 23nd annual ACM symposium on user interface software and technology (pp. 303–312). New York, NY, USA: ACM.

Chen, J., Nairn, R., Nelson, L., Bernstein, M., & Chi, E. (2010). Short and tweet: experiments on recommending content from information streams. In E. Mynatt, G. Fitzpatrick, S. Hudson, K. Edwards, & T. Rodden (Eds), Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1185–1194). New York, NY, USA: ACM.

Chen, K., Chen, T., Zheng, G., Jin, O., Yao, E., & Yu, Y. (2012). Collaborative personalized tweet recommendation. In W. Hersh, J. Callan, Y. Maarek, & M. Sanderson (Eds), Proceedings of the 35th international ACM SIGIR conference on research and development in information retrieval (pp. 661–670). New York, NY, USA: ACM.

Douglis, F. (2009). Information overload, 140 characters at a time. IEEE Internet Computing, 13(4), 4–5.

Finin, T., Murnane, W., Karandikar, A., Keller, N., Martineau, J., & Dredze, M. (2010). Annotating named entities in Twitter data with crowdsourcing. In C. Callison-Burch, & M. Dredze (Eds), Proceedings of the NAACL HLT 2010 workshop on creating speech and language data with amazon's mechanical turk (pp. 80–88). Stroudsburg, PA, USA: Association for Computational Linguistics.

Grineva, M., & Grinev, M. (2012). Information overload in social media streams and the approaches to solve it. WWW-2012.

Guo, Y., Goh, D.H.-L., Ilangovan, K., Jiao, S., & Yang, X. (2012). Investigating factors influencing non-use and abandonment of microblogging services. Journal of Digital Information Management, 10(6), 421.

Hargittai, E., Neuman, W.R., & Curry, O. (2012). Taming the information tide: Perceptions of information overload in the American home. The Information Society, 28(3), 161–173.

Haveliwala, T., & Kamvar, S. (2003). The second eigenvalue of the google matrix. Stanford University Technical Report.

Haveliwala, T.H. (2003). Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search. IEEE Transactions on Knowledge and Data Engineering, 15(4), 784–796.

Heaivilin, N., Gerbert, B., Page, J., & Gibbs, J. (2011). Public health surveillance of dental pain via Twitter. Journal of Dental Research, 90(9), 1047–1051.

Inouye, D., & Kalita, J.K. (2011). Comparing Twitter summarization algorithms for multiple post summaries. In A. (Sandy) Pentland, J. Clippinger, & L. Sweeney (Eds), Privacy, security, risk and trust (passat), 2011 IEEE third international conference on and 2011 IEEE third international conference on social computing (socialcom) (pp. 298–306). Washington, DC, USA: IEEE Computer Society.

Lage, R., Durao, F., & Dolog, P. (2012). Towards effective group recommendations for microblogging users Proceedings of the 27th Annual ACM Symposium on Applied Computing, 923–928. ACM.

Lage, R., Dolog, P., & Leginus, M. (2014a). The role of adaptive elements in web-based surveillance system user interfaces. In V. Dimitrova, T. Kuflik, D. Chin, F. Ricci, P. Dolog, & G.-J. Houben (Eds.), User modeling, adaptation, and personalization (Vol. 8538, pp. 350–362). Berlin: Springer International Publishing.

Lage, R., Dolog, P., & Leginus, M. (2014b). Vector space models for the classification of short messages on social network services. In K.-H. Krempels & A. Stocker (Eds.), Web information systems and technologies (pp. 209–224). Berlin Heidelberg: Springer.

Leginus, M., Dolog, P., & Lage, R. (2013). Graph based techniques for tag cloud generation. In G. Stumme, & A. Hotho (Eds), Proceedings of the 24th ACM conference on hypertext and social media (pp. 148–157). New York, NY, USA: ACM.

Mei, Q., Guo, J., & Radev, D. (2010). Divrank: the interplay of prestige and diversity in information networks. In B. Rao, B. Krishnapuram, A. Tomkins, & Q. Yang (Eds), Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1009–1018). New York, NY, USA: ACM.

Michelson, M., & Macskassy, S.A. (2010). Discovering users' topics of interest on Twitter: a first look. In R. Basili, D. Lopresti, C. Ringlstetter, S. Roy, K.U. Schulz, & L.V. Subramaniam (Eds.), Proceedings of the fourth workshop on analytics for noisy unstructured text data (pp. 73–80). New York, NY, USA: ACM.

Mihalcea, R., & Tarau, P. (2004). Textrank: Bringing order into texts. In D. Lin & D. Wu (Eds), Proceedings of EMNLP (pp. 404–411). Barcelona, Spain: Association for Computational Linguistics.

Naaman, M., Boase, J., & Lai, C.-H. (2010). Is it really about me?: message content in social awareness streams. In K. Inkpen, C. Gutwin, & J. Tang (Eds), Proceedings of the 2010 ACM conference on computer supported cooperative work (pp. 189–192). New York, NY, USA: ACM.

Nguyen, T.T., Hui, P.-M., Harper, F.M., Terveen, L., & Konstan, J.A. (2014). Exploring the filter bubble: the effect of using recommender systems on content diversity. In C.-W. Chung, A. Broder, K. Shim, & T. Suel (Eds), Proceedings of the 23rd international conference on world wide web (pp. 677–686). New York, NY, USA: ACM.

O'Connor, B., Krieger, M., & Ahn, D. (2010). Tweetmotif: Exploratory search and topic summarization for Twitter. In Icwsm.

Qu, Z., & Liu, Y. (2011). Interactive group suggesting for Twitter. In D. Lin (Ed.), ACL (short papers) (pp. 519–523). Stroudsburg, PA, USA: Association for Computational Linguistics.

Rout, D., Bontcheva, K., & Hepple, M. (2013). Reliably evaluating summaries of Twitter timelines. In Proceedings of the AAAI workshop on analyzing microtext, Palo Alto, California: AAAI.

Shen, X., Tan, B., & Zhai, C. (2005). Context-sensitive information retrieval using implicit feedback. In R. Baeza-Yates, N. Ziviani, G. Marchionini, A. Moffat, & J. Tait (Eds), Proceedings of the 28th annual international ACM SIGIR conference on research and development in information retrieval (pp. 43–50). New York, NY, USA: ACM.

Soboroff, I., Ounis, I., Lin, J., & Soboroff, I. (2012). Overview of the trec-2012 microblog track. In E.M. Voorhees, & L.P. Buckland (Eds), Proceedings of the twenty-first text retrieval conference (TREC 2012). Gaithersburg, Maryland: National Institute of Standards and Technology.

Uysal, I., & Croft, W.B. (2011). User oriented tweet ranking: A filtering approach to microblogs. In B. Berendt, A. de Vries, W. Fan, C. Macdonald, I. Ounis, & I. Ruthven (Eds), Proceedings of the 20th ACM international conference on information and knowledge management (pp. 2261–2264). New York, NY, USA: ACM.

Venetis, P., Koutrika, G., & Garcia-Molina, H. (2011). On the selection of tags for tag clouds. In I. King, W. Nejdl, & H. Li (Eds), Proceedings of the fourth ACM international conference on web search and data mining (pp. 835–844). New York, NY, USA: ACM.

White, S., & Smyth, P. (2003). Algorithms for estimating relative importance in networks. In L. Getoor, T. Senator, P. Domingos, & C. Faloutsos (Eds), Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining (pp. 266–275). New York, NY, USA: ACM.

Wu, W., Zhang, B., & Ostendorf, M. (2010). Automatic generation of personalized annotation tags for Twitter users. In R.M. Kaplan (Ed.), Human language technologies: The 2010 annual conference of the North American chapter of the Association for Computational Linguistics (pp. 689–692). Stroudsburg, PA, USA: Association for Computational Linguistics.

Yan, R., Lapata, M., & Li, X. (2012). Tweet recommendation with graph co-ranking. In H. Li, C.-Y. Lin, & M. Osborne (Eds), Proceedings of the 50th annual meeting of the association for Computational Linguistics: Long papers (Vol. 1, pp. 516–525). Stroudsburg, PA, USA: Association for Computational Linguistics.

Yu, J., Shen, Y., & Xie, J. (2013). Mining user interest and its evolution for recommendation on the micro-blogging system. In J. Wang, H. Xiong, Y. Ishikawa, J. Xu, & J. Zhou (Eds), Web-age information management (pp. 679–690). Berlin, Heidelberg: Springer.